Document made available under the Patent Cooperation Treaty (PCT)

International application number: PCT/US05/010436

International filing date: 31 March 2005 (31.03.2005)

Document type: Certified copy of priority document

Document details: Country/Office: US
Number: 60/558,862
Filing date: 02 April 2004 (02.04.2004)

Date of receipt at the International Bureau: 02 May 2005 (02.05.2005)

Remark: Priority document submitted or transmitted to the International Bureau in compliance with Rule 17.1(a) or (b)

1311486

THE UNITED STATES OF AMERICA

TO ALL TO WHOM THESE PRESENTS SHALL COME:

UNITED STATES DEPARTMENT OF COMMERCE

United States Patent and Trademark Office

April 20, 2005

THIS IS TO CERTIFY THAT ANNEXED HERETO IS A TRUE COPY FROM THE RECORDS OF THE UNITED STATES PATENT AND TRADEMARK OFFICE OF THOSE PAPERS OF THE BELOW IDENTIFIED PATENT APPLICATION THAT MET THE REQUIREMENTS TO BE GRANTED A FILING DATE.

APPLICATION NUMBER: ***60/558,862***
FILING DATE: ***April 02, 2004***
RELATED PCT APPLICATION NUMBER: ***PCT/US05/10436***

UNITED STATES PATENT AND TRADEMARK OFFICE
DEPARTMENT OF COMMERCE

Certified by

Jon W. Dudas

Under Secretary of Commerce
for Intellectual Property
and Director of the United States
Patent and Trademark Office

16569 U.S.PTO 040204

Please type a plus sign (+) inside this box → +

Express Mail Label: EV441452350US

PTO/SB/16 (8-00)
Approved for use through 10/31/2002. OMB 0651-0032
Patent and Trademark Office; U.S. DEPARTMENT OF COMMERCE

15535 U.S. PTO 60/558862 040204

PROVISIONAL APPLICATION FOR PATENT COVER SHEET

This is a request for filing a PROVISIONAL APPLICATION FOR PATENT under 37 CFR 1.53 (c).

INVENTOR(S)

Given Name (first and middle [if any])	Family Name or Surname	Residence (City and either State or Foreign Country)
Jill MacDonald	Boyce	Manalapan, New Jersey
Alexandros	Tourapis	West Windsor, New Jersey

☐ *Additional inventors are being named on the _____ separately numbered sheets attached hereto*

TITLE OF THE INVENTION (280 characters max)

METHOD AND APPARATUS FOR COMPLEXITY SCALABLE CODEC

CORRESPONDENCE ADDRESS

Direct all correspondence to:

☐ Customer Number [ ] → *Place Customer Number Bar Code Label here*

OR *Type Customer Number here*

☒ Firm *or* Individual Name	Joseph S. Tripoli - Thomson Licensing Inc.				
Address	PATENT OPERATIONS				
Address	P. O. BOX 5312				
City	PRINCETON	State	NJ	ZIP	08543-5312
Country	USA	Telephone	609-734-6834	Fax	609-734-6888

ENCLOSED APPLICATION PARTS *(check all that apply)*

☒ Specification *Number of Pages* 9 ☐ CD(s), Number

☒ Drawing(s) *Number of Sheets* 6 ☐ Other (specify)

☐ Application Data Sheet. See 37 CFR 1.76

METHOD OF PAYMENT OF FILING FEES FOR THIS PROVISIONAL APPLICATION FOR PATENT *(check one)*

☐ Applicant claims small entity status. See 37 CFR 1.27.

☐ A check or money order is enclosed to cover the filing fees

☒ The Commissioner is hereby authorized to charge filing fees or credit any overpayment to Deposit Account Number: 07-0832

FILING FEE AMOUNT ($): 160

☐ Payment by credit card. Form PTO-2038 is attached.

The invention was made by an agency of the United States Government or under a contract with an agency of the United States Government.

☒ No.

☐ Yes, the name of the U.S. Government agency and the Government contract number are: _____.

Respectfully submitted,

SIGNATURE [signature]

Date 04/02/04

TYPED or PRINTED NAME GUY H. ERIKSEN

REGISTRATION NO. (*if appropriate*) 41,736

TELEPHONE (609) 734-6807

Docket Number: PU040098

USE ONLY FOR FILING A PROVISIONAL APPLICATION FOR PATENT

METHOD AND APPARATUS FOR COMPLEXITY SCALABLE CODEC

The present invention is directed towards video coders and/or decoders (CODECs), and more particularly towards an apparatus and method for scalable complexity CODECs.

It is desirable for a broadcast video application to provide support for diverse user devices, without incurring the bitrate penalty associated with simulcast encoding. Video decoding is a complex operation, and the complexity is very dependent on the resolution of the coded video. Low power portable devices typically have very strict complexity restrictions and low resolution displays. Simulcast broadcast of two or more video bitstreams corresponding to different resolutions can be used to address the complexity requirements of the lower resolution devices, but requires a higher total bitrate than a complexity scalable system of this invention. This invention provides a solution that allows for complexity scalable decoders while maintaining high video coding bitrate efficiency.

Many different methods of scalability have been widely studied and standardized, including SNR scalability, spatial scalability, temporal scalability, and fine grain scalability, in scalability profiles of the MPEG-2 and MPEG-4 standards. [1], [2], [4]. Most of the work in scalable coding has been aimed at bitrate scalability, where the low resolution layer has a limited bandwidth. Figure 1 shows a typical spatial scalability system, where low resolution decoders are connected to a low bandwidth network, and high resolution decoders are connected to a high bandwidth network. Scalable coding has not been widely adopted in practice, because of the considerable increase in encoder and decoder complexity, and because the coding efficiency of scalable encoders is typically well below that of non-scalable encoders.

Spatially scalable encoders and decoders typically require that the high resolution scalable encoder/decoder provide additional functionality than would be present in a normal high resolution encoder/decoder. In an MPEG-2 spatial scalable encoder, a decision is made whether prediction is performed from a low resolution picture or from a high resolution reference picture. An MPEG-2 spatial scalable decoder must be capable of predicting either from the low resolution picture or the

high resolution picture. Two sets of reference picture stores are required by an MPEG-2 spatial scalable encoder/decoder, one for low resolution pictures and another for high resolution pictures. Figure 2 shows a block diagram of a spatial scalable encoder supporting two layers. Figure 3 shows a block diagram of a spatial scalable decoder supporting two layers.

In Figure 2, a high resolution input video sequence is received. It is downsampled to create a low resolution video sequence. The low resolution video sequence is encoded using a normal low resolution video compression encoder, creating a low resolution bitstream. The low resolution bitstream is decoded using a normal low resolution video compression decoder. (This function may be performed inside of the encoder.) The decoded low resolution sequence is upsampled, and provided as one of two inputs to a scalable high resolution encoder. The scalable high resolution encoder encodes the video to create a high resolution scalable bitstream.

In Figure 3, both a high resolution scalable bitstream and low resolution bitstream are received. The low resolution bitstream is decoded using a normal low resolution video compression decoder, which utilizes low resolution frame stores. The decoded low resolution video is upsampled, and then input into a high resolution scalable decoder. The high resolution scalable decoder utilizes a set of high resolution frame stores, and creates the high resolution output video sequence.

Figure 4 shows a block diagram of a typical non-scalable video encoder used in the H.264/MPEG AVC standard.[3] Figure 4 shows a block diagram of a typical non-scalable video decoder used with H.264/MPEG AVC. Figure 5 shows a block diagram of a normal non-scalable video decoder. In an earlier filed provisional patent application [5], it was proposed that H.264/MPEG AVC be extended to use a Reduced Resolution Update (RRU) mode. The RRU mode improves coding efficiency at low bitrates by reducing the number of residual MBs to be coded, while performing motion estimation and compensation of full resolution pictures. Figure 6 shows a RRU video encoder. Figure 7 shows a RRU video decoder.

[1] MPEG-2, ISO/IEC 12818-2, "Generic coding of moving pictures and associated audio information: Video"
[2] MPEG-4, 14496-2:1999, "Coding of audio-visual objects"
[3] Wiegand, "Draft ITU-T Recommendation and Final Draft International Standard of Joint Video Specification (ITU-T Rec. H.264 I ISO/IEC 14496-10 AVC)", Mar 31, 2003.
[4] F. Wu, S. Li, R. Yan, X. Sun, and Y. Zhang, "Efficient and Universal Scalable Video Coding," ICIP 2002.
[5] A. Tourapis and J. Boyce, " Reduced Resolution Slice Update Mode for Advanced Video Coding," PU040073, U.S. Provisional Patent Application No. 60/551,417 filed on March 9, 2004.

The present invention is useful in that it enables a broadcast video system with diverse user endpoint devices, while maintaining coding efficiency. Without loss in generality, consider a system which supports two different levels of decoder complexity and resolution. A low resolution decoder has a smaller display size and has very strict decoder complexity constraints. A full resolution decoder has a larger display size and less strict but still important decoder complexity constraints.

A broadcast or multicast system transmits two bitstreams, a base layer with bitrate BR_{base} and an enhancement layer with bitrate BR_{enhan}. The two bitstreams may be multiplexed together and sent in a single transport stream. Figure 8 illustrates a complexity scalability broadcast system, which includes a complexity scalability video encoder and a low resolution decoder and a full resolution decoder. The low resolution decoder processes only the base layer bitstream and the full resolution decoder processes both the base layer bitstream and the enhancement layer bitstream.

A key goal of this system is to minimize $BR_{base} + BR_{enhan}$. This differs somewhat from a typical scalability system where minimizing BR_{base} itself is also considered important, as shown in Figure 1 where the low resolution devices are connected via low bandwidth network. In the complexity scalability system, it is assumed that both the base layer and the enhancement layer are broadcast, so the bitrate of the base layer bitstream itself is not necessarily as highly constrained.

In accordance with the principles of the present invention, the bits used for coding of the video residual formed after motion estimation/compensation are used in both the low resolution decoder and the full resolution decoder. The motion vectors (mvs) transmitted in the base layer bitstream are used in both the low resolution decoder and the full resolution decoder, but with a higher accuracy in the

full resolution decoder than in the low resolution decoder. Also, the motion compensation prediction is done at a low resolution in the low resolution decoder and at a high resolution in the high resolution decoder. Similarly to what is done in the RRU codec of Figures 6 and 7, the motion blocks at the low resolution correspond to larger blocks at the high resolution. So, when applied to the H.264/MPEG AVC codec, for example, the allowable motion block sizes of 16x16, 16x8, 8x16, 8x8, 8x4, 4x8 and 4x4 are used in the low resolution base layer, but correspond to larger block sizes of 32x32, 32x16, 16x32, 16x16, 16x8, 8x16, and 8x8, respectively, at the full resolution.

The low resolution decoder uses only the base layer bitstream. An additional enhancement layer bitstream is also transmitted, e.g. using 16x16 macroblocks, for use in the full resolution decoder. The enhancement layer bitstream includes a full resolution error signal, to be added to the result of decoding of the base layer bitstream, which was done with full resolution motion compensation. The bitrate of the enhancement layer may end up being lower than that of the base layer, which differs from the typical spatial scalability case where the base layer bitrate is typically small compared with the enhancement layer bitrate. A full resolution error signal is not necessarily sent for every coded macroblock or slice/picture.

Figure 9 shows a block diagram of a low resolution decoder in accordance with the principles of the present invention. The base layer bitstream is entropy decoded. The motion vectors are rounded to reduce them in accuracy to correspond to the low resolution. The remaining blocks are identical to those found in a standard video decoder, including inverse quantization and inverse transform, motion compensation, and deblocking filter. The complexity of this low resolution scalable decoder is very similar to that of a non-scalable decoder, as scaling of motion vectors is of very low complexity. If factors of 2 are used in the resolution ratios in each dimension between the low and full resolution, the rounding can be implemented with just a right shift or an add and a right shift, depending whether rounding up or rounding down is selected in the system.

In an alternative embodiment of the present invention, the motion vectors transmitted in the base layer are not of the higher resolution. In this case, the low resolution decoder can be completely backwards compatible with an existing coding standard. However, such a system may be of lower coding efficiency as the additional bit accuracy of the motion vectors for the full resolution are transmitted in

the enhancement layer bitstream. In this case the enhancement layer could be coded similar to a P slice, and motion vectors are differentially coded first based on layer prediction (i.e. differentially coded versus the corresponding low resolution layer mv), and secondly using spatial prediction (i.e. differentially coded versus adjacent mvs or even versus adjacent differential mvs).

Figure 10 shows a block diagram of a full resolution decoder in accordance with the present invention. The portion of the decoder that operates on the base layer bitstream is similar to an RRU decoder. After entropy decoding and inverse quantization and inverse transform, the residual is upsampled. Motion compensation is applied to the full resolution reference pictures to form a full resolution prediction, and the upsampled residual is added to the prediction. If a full resolution error signal is present in the enhancement layer bitstream, it is entropy decoded and inversed quantized and transformed, and then added to the RRU reconstructed signal. The deblocking filter is then applied. Presence of full resolution error signal could be signaled at the macroblock level with the use of a *Skip* macroblock mode. If a macroblock is marked as skipped no additional error signal is present, while if not, the delta_quant, the coded block pattern and the actual residual have to also be transmitted. Skip macroblocks could also be run-length coded to further increase efficiency. An additional intra directional prediction mode may be created that performs no directional prediction. Although it may be more efficient to not perform any additional prediction if a macroblock in the enhancement layer is skipped, additional prediction could also be inferred by considering adjacent macroblocks. For example, if all intra prediction modes as described in H.264 are available, then an additional prediction for skip could also be generated which can be derived from the prediction modes of the adjacent macroblocks (i.e. minimum directional prediction) which is then added to the RRU reconstructed signal to generate the final prediction. Similarly, an additional direct intra mode could also be used which could also derive its directional prediction mode from adjacent macroblocks, while still allowing the transmission of an error signal.

A key difference in this architecture from a traditional spatial scalable decoder is that there is no need for two sets of reference pictures stores and motion compensation units. This full resolution decoder contains only full resolution reference pictures stores and only performs motion compensation once at the full resolution. In contrast, the spatial scalability decoder of Figure 3 includes both full resolution and

low resolution reference pictures stores, and performs motion compensation at both the full resolution and the low resolution. This leads to a significant reduction in computations, memory, and memory bandwidth for full resolution decoders in accordance with this invention as compared to traditional spatial scalable decoders.

The decoder complexity of the full resolution scalable decoder is similar to that of a normal video decoder of the same resolution. The inverse quantization and inverse transform blocks for the base layer bitstream are of lower complexity, as they operate on few blocks that a normal decoder. However, additional entropy decoding and inverse quantization and inverse transform are used for the enhancement layer bitstream. The motion compensation and the deblocking filter, which are the most computationally complex blocks of a decoder, are unchanged from a normal decoder.

In an embodiment of the present invention, an enhancement layer bitstream full resolution error signal is only sent when intra-coded (I) slices are present in the base layer. Limiting the use of the enhancement layer for only I slices limits the decoder complexity for software implementations. I slices generally require fewer computations than P and B slices, and hence there should be spare CPU cycles available for the additional entropy decode and inverse quantization and inverse transform operations.

Figure 11 shows an example of a Complexity Scalable Video Encoder. This encoder attempts to optimize the full resolution video quality rather than the low resolution video quality. Motion estimation is performed on the full resolution video picture. After subtraction the motion compensated prediction from the input picture, the prediction residual is downsampled. Unlike in the RRU codec, the downsampling is applied to all pictures, so that the low resolution decoder can always have a picture to decode. The downsampled residual is transformed and quantized, and entropy coded. This forms the base layer bitstream. The inverse quantizer and inverse transform is applied, and then the coded residual is upsampled back to the full resolution. The encoder can choose whether or not to send an enhancement layer full resolution error signal for the picture or slice. In general, an enhancement layer full resolution error signal is coded for all I slices, and can be optionally sent for P and B slices based on the magnitude of the error signal when the full resolution input picture subtracts the decoded upsampled. If an enhancement layer full resolution error signal is to be coded, the coded base layer upsampled coded picture is

subtracted from the input full resolution picture. The difference is then quantized, transformed and entropy coded to form the enhancement layer bitstream. The enhancement layer bitstream can be seen as containing only intra-coded slices.

In an alternative embodiment, a joint optimization of both the low resolution and full resolution pictures could take place. That would require addition of a full low resolution decoder model inside of the scalable encoder, and low resolution reference pictures stores, and an additional low resolution motion estimation block. Any of several different upsampling and downsampling filters can be used, for example bilinear interpolation, zero order hold, or multi-tap filters.

Additional deblocking filters could be added in the full resolution decoder and the scalable encoder, prior to the addition of the enhancement layer error signal. Deblocking could in this case also consider the enhancement layer macroblock modes used, i.e. if all affected blocks are skipped, no additional deblocking is applied, otherwise different strength filtering is applied depending on whether the upscaling was performed on the residual or the low resolution reconstructed block.

There is more than one possible method to use for intra prediction in the full resolution decoder, when applied to H.264/MPEG AVC. Intra prediction could be applied at the low resolution, using the same prediction pixels as in the H.264/MPEG AVC spec. Alternatively, the method described in U.S. Provisional Patent Application No. 60/551,417 (Docket Number PU040073), filed on March 9, 2004, could be used, in which the intra prediction is applied at the full resolution, and a larger number of pixels at the full resolution are used in the prediction.

In an alternative embodiment, the full resolution decoder may decide performing motion compensation for a macroblock using the same resolution and method as for the base layer decoding (i.e. using 16x16 macroblocks), which is then upsampled to full resolution. Upsampling could be performed using a bilinear or longer tap filter. A full resolution error signal could also be added, if available. The decision could be made through additional signaling at the macroblock level (i.e. with the presence of a *RRU* macroblock mode, and a *low resolution macroblock* mode, apart from *SKIP* mode). This process may be desirable for certain cases where, due to high motion and texture detail, upsampling the residual would lead to the generation of undesirable high frequencies and artifacts. Nevertheless, this would also require that the full resolution decoder is able to store, or on the fly generate, the low resolution references. The longer tap filter could also incur further

complexity, which is although partly compensated from the fact that motion compensation is performed for a smaller macroblock. A second, simpler alternative solution to the same problem however is to perform motion compensation at full resolution, entropy decode and inverse quantize and inverse transform the base layer residual but not add it to the motion compensated signal, prior to finally adding the full resolution error. This method requires decoding of the base layer residual in order to update the entropy context model for the decoding of the remaining residuals. This later solution could replace the low resolution macroblock mode, or co-exist as an additional mode for the encoding of the full resolution residual.
The above description and figures assume two layers of scalability, however, this concept can be extended to an arbitrary number of layers.

Advantages and Advancements

1. Video decoder in which decoded motion vectors are reduced in accuracy while not reducing the resolution of the reference picture prediction.
2. Scalable video decoder where a base layer prediction residual is upsampled, added to the full resolution motion compensated prediction, and added to a decoded full resolution enhancement layer error signal.
3. Decoder of item 2 in which enhancement layer error signal is intra coded.
4. Application of a deblocking filter to the decoder of item 2, following the addition of the error signal.
5. Deblocking filter of item 4 that also considers the enhancement layer mode signals.
6. Scalable video encoder which creates a base layer and an enhancement layer, with the enhancement layer created only for intra-coded slices in the base layer.
7. Scalable video encoder which creates a base layer and an enhancement layer, the base layer containing a coded low resolution prediction residual and full resolution motion vectors, and the enhancement layer containing a coded error signal formed by subtracting from the input picture the result of the decoder in item 2.
8. Scalable video decoder containing a full resolution reference picture store and full resolution motion compensation block, a lower resolution inverse quantizer and inverse transform, and an upsampler.

9. Scalable video decoder which processes both a base layer bitstream and an enhancement layer bitstream but includes only full resolution reference picture stores and not low resolution reference picture stores.
10. Decoder of item 2 where optionally, i.e. based on a enhancement layer signal, the base layer prediction residual is not added to the full resolution motion compensated prediction. This is then added to the decoded full resolution enhancement layer error signal.
11. Decoder of item 2 where optionally, i.e. again based on a enhancement layer signal, the base layer prediction residual is added to the low resolution motion compensated prediction. This is then upsampled and added to the decoded full resolution enhancement layer error signal.
12. Decoder of item 2 where enhancement layer error signal contains skipped residual macroblocks, and macroblock modes signaling how prediction signal will be generated based on items 3, 10, and 11.

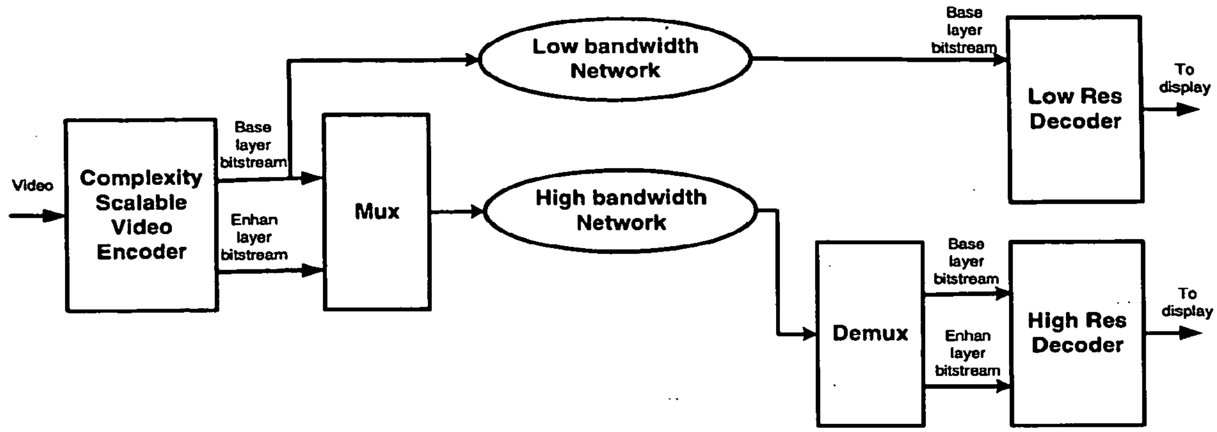

Figure 1. Typical Spatial Scalability System

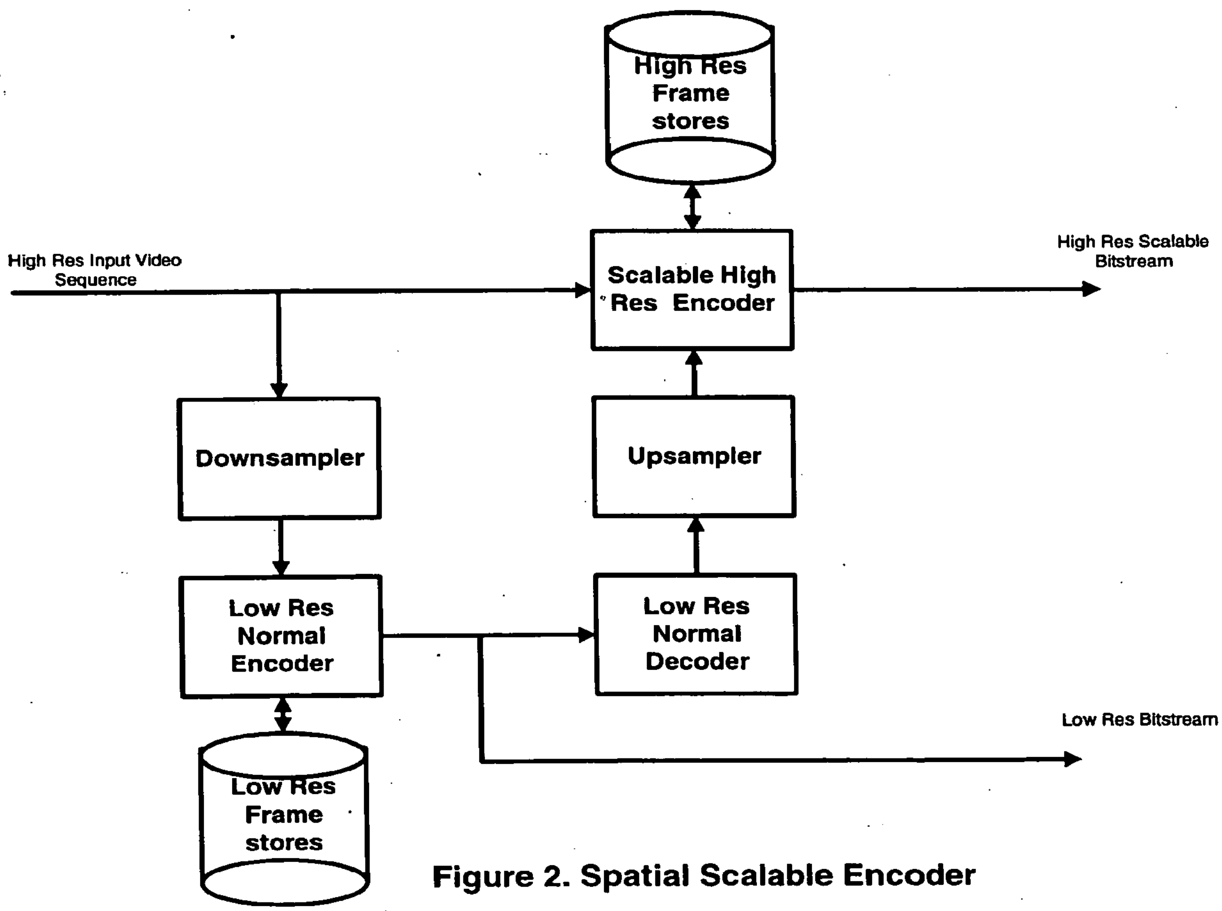

Figure 2. Spatial Scalable Encoder

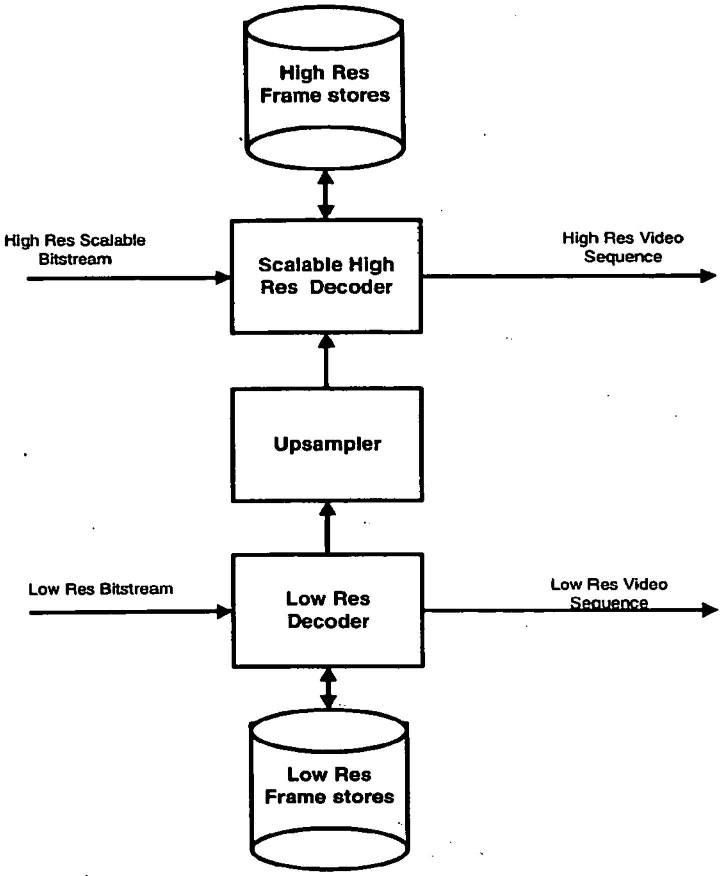

Figure 3. Spatial Scalable Decoder

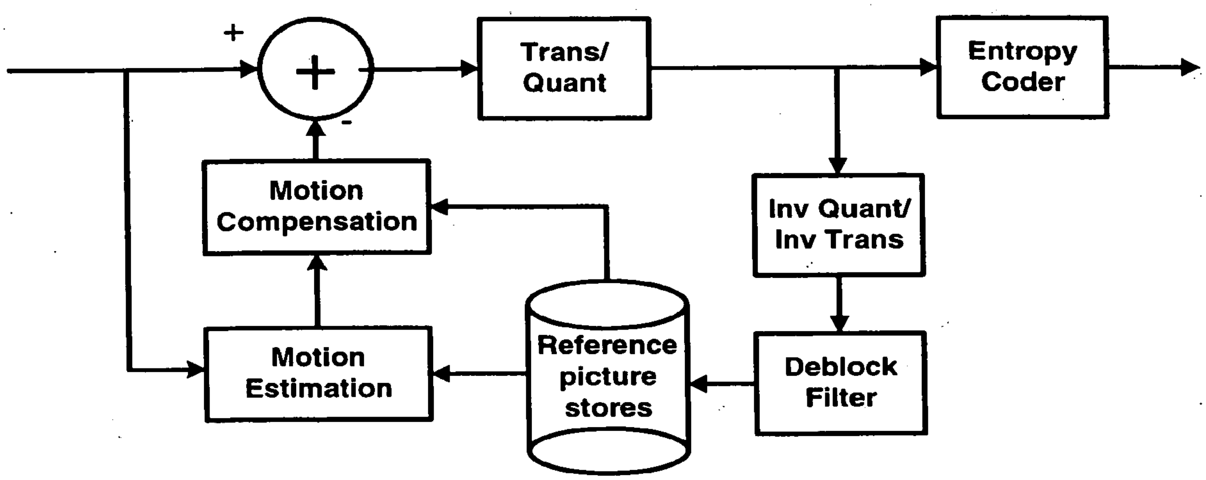

Figure 4. Standard Video Encoder

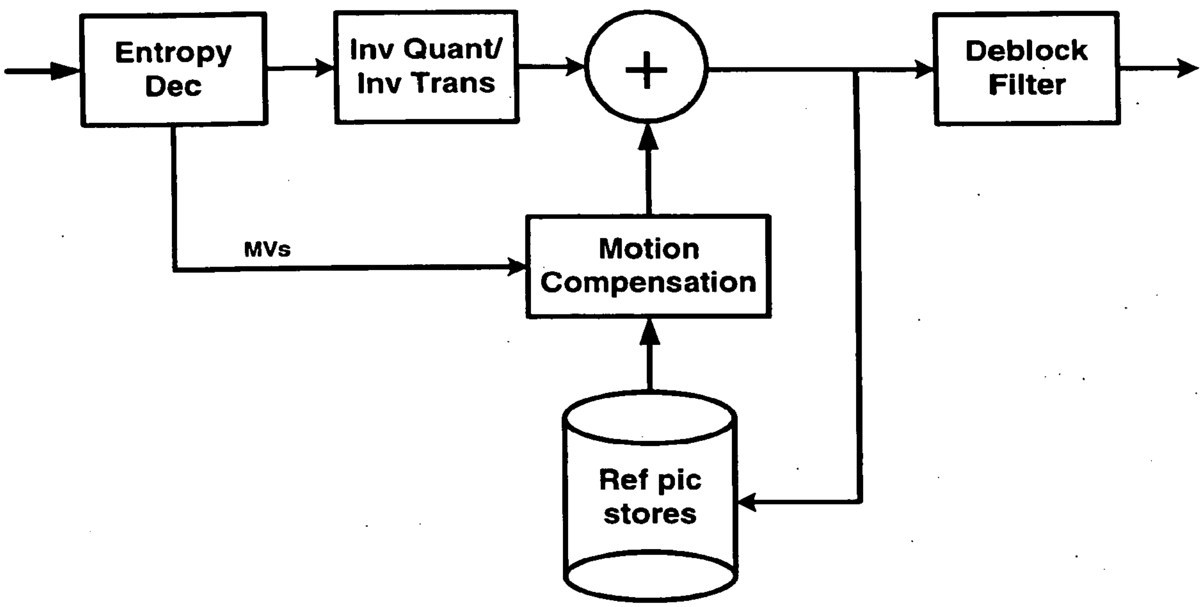

Figure 5. Standard Video Decoder

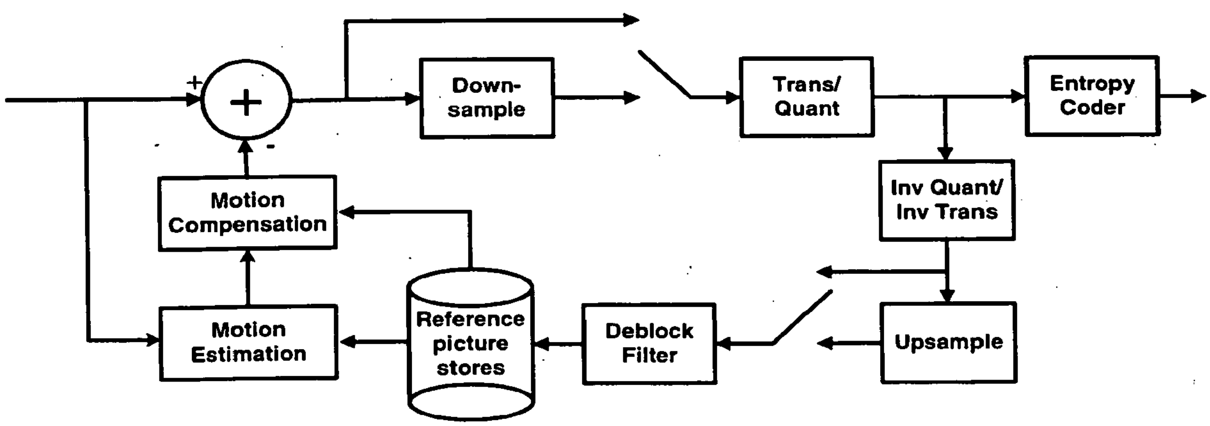

Figure 6. RRU Video Encoder

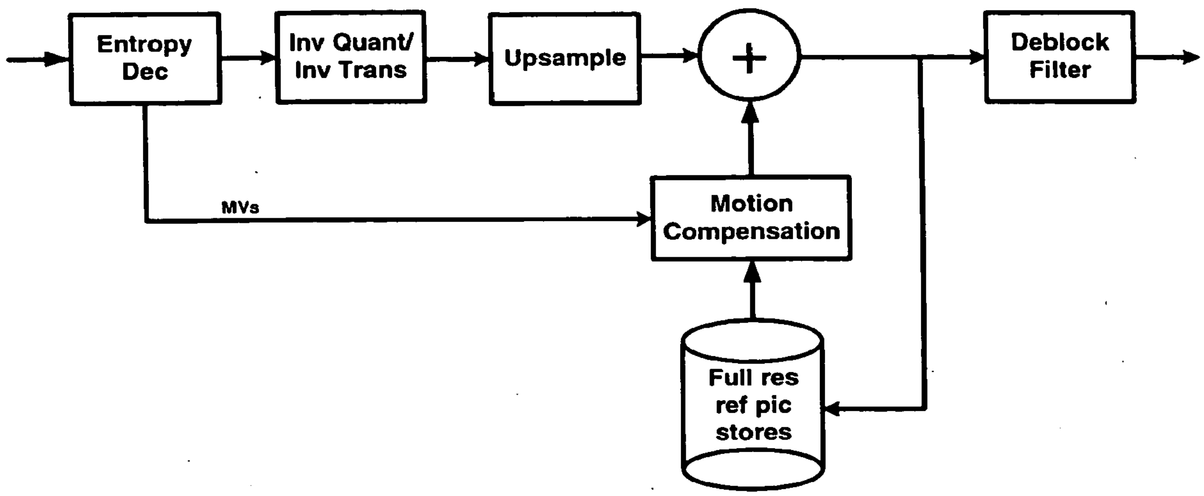

Figure 7. RRU Video Decoder

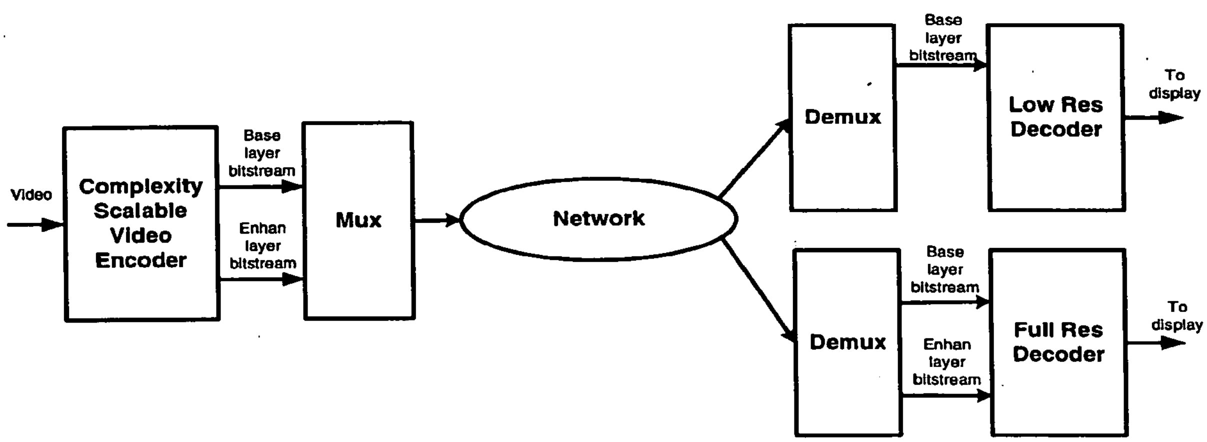

Figure 8. Complexity Scalability Broadcast System

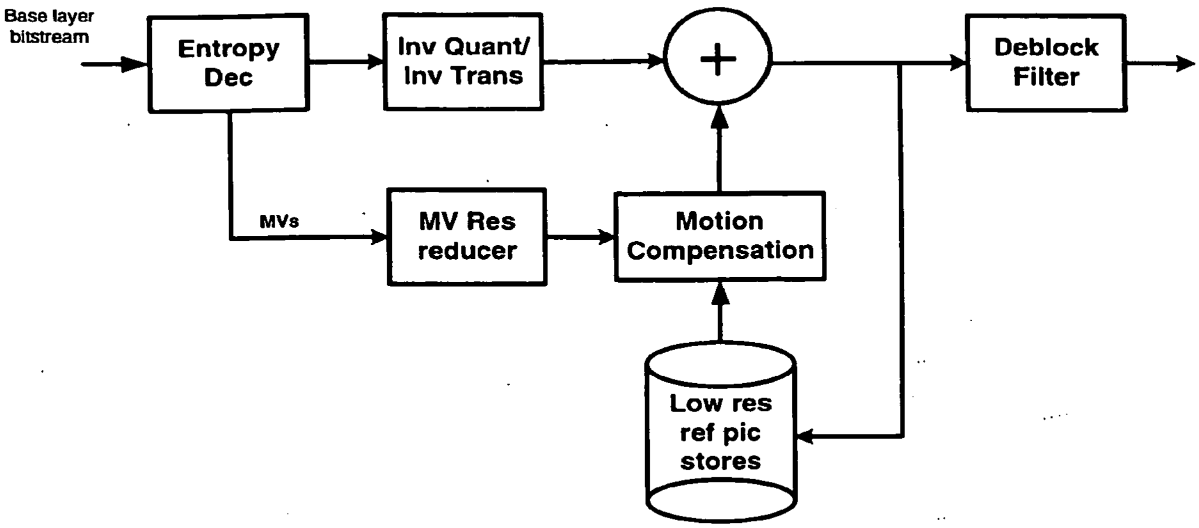

Figure 9. Low Res Complexity Scalable Decoder

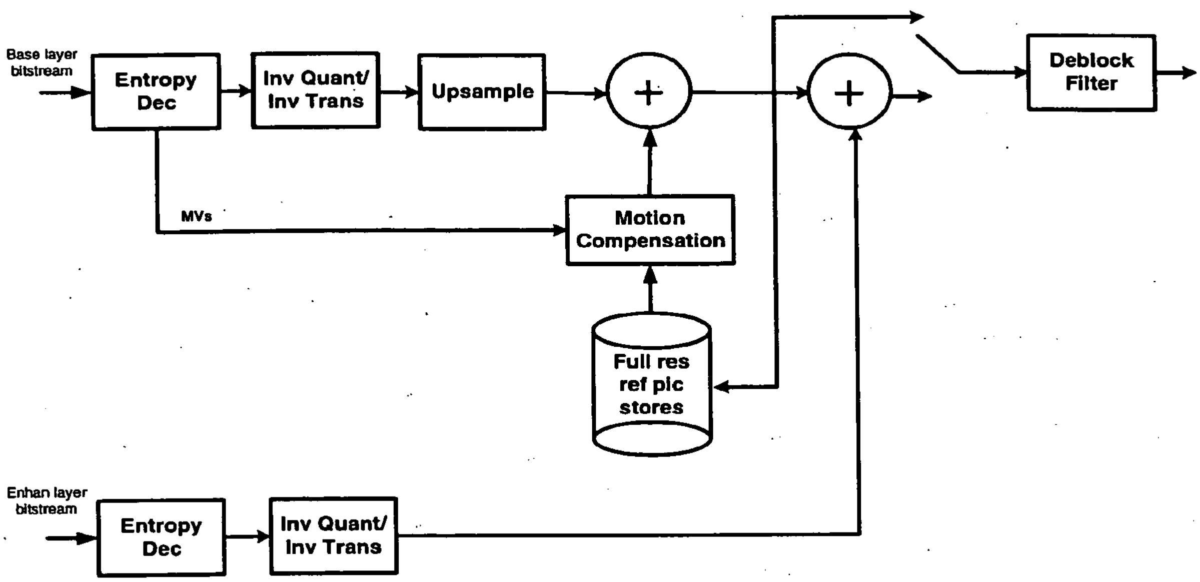

Figure 10. High Res Complexity Scalable Decoder

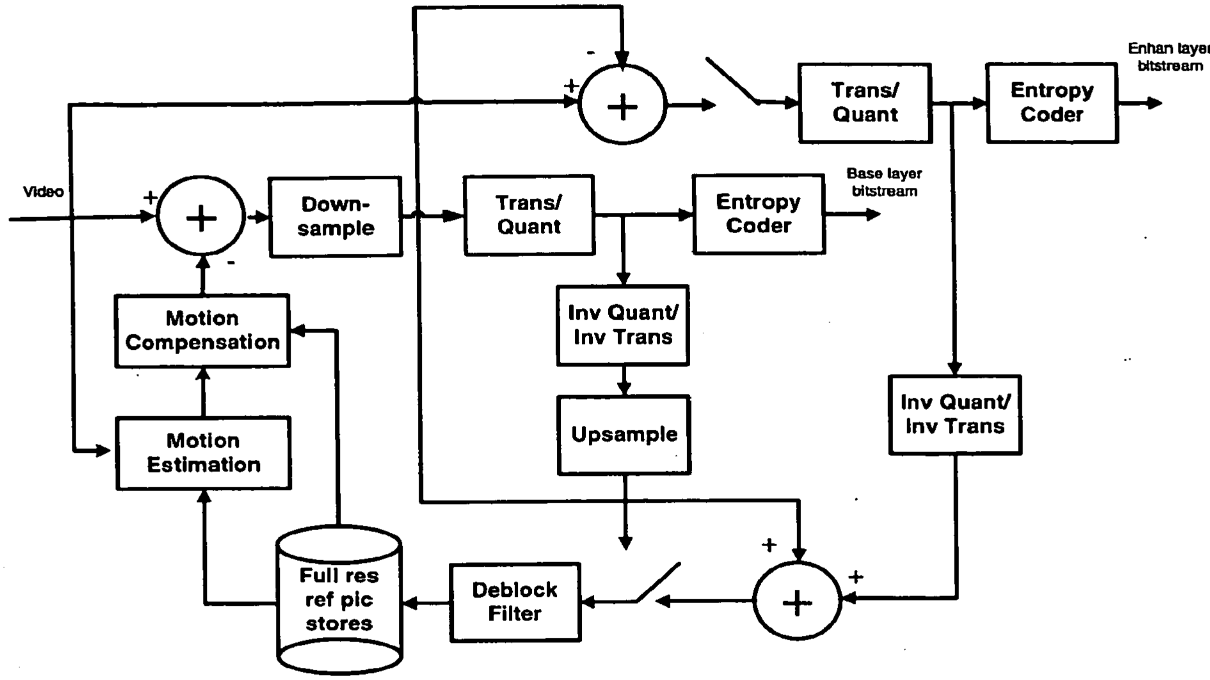

Figure 11. Complexity Scalable Video Encoder